AFRL-HE-AZ-TR-2006-0015-Vol V

Air Force Research Laboratory

Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Using the Pathfinder Methodology to Assess Pilot Knowledge Structure Changes

Brian T. Schreiber
Pam DiSalvo
William A. Stock

Lumir Research Institute
195 Bluff Avenue
Grayslake IL 60030

and

Winston Bennett, Jr.

Air Force Research Laboratory
Warfighter Readiness Research Division
6030 South Kent Street
Mesa AZ 85212-6061

Volume V

July 2006

Final Report for March 2002 to July 2005

Human Effectiveness Directorate
Warfighter Readiness Research Division
6030 South Kent Street
Mesa, AZ 85212-6061

NOTICES

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-AZ-TR-2006-0015-Vol V

This technical report has been reviewed and is approved for publication.

// Signed //

WINSTON BENNETT JR.
Project Scientist

// Signed //

HERBERT H. BELL
Technical Advisor

// Signed //

DANIEL R. WALKER, Colonel, USAF
Chief, Warfighter Readiness Research Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
03/31/2006	Final	March 2002 to July 2005

4. TITLE AND SUBTITLE

Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Using the Pathfinder Methodology to Assess Pilot Knowledge Structure Changes

5a. CONTRACT NUMBER F41624-97-D-5000

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER 62202F

6. AUTHOR(S)

Schreiber, Brian T.[1]; DiSalvo, Pam[1], Stock, William A. [1], Bennett, Winston Jr.[2]

5d. PROJECT NUMBER 1123

5e. TASK NUMBER AS

5f. WORK UNIT NUMBER 03

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

1. Lumir Research Institute
195 Bluff Ave
Grayslake, IL 60030

2. Air Force Research Laboratory
Warftr Readiness Research Div
6030 South Kent Street
Mesa, AZ 85212-6061

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

10. SPONSOR/MONITOR'S ACRONYM(S) AFRL; AFRL/HEA

11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-HE-AZ-TR-2006-0015 – Vol V

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
This is the last volume of a five-volume report, Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study.

14. ABSTRACT

To provide an indirect assessment of learning for the overall Distributed Mission Operations (DMO) Within Simulator Training Effectiveness Baseline Study as described in Volume I, Summary Report, of AFRL-HE-AZ-TR-2006-0015, the current work examined pilots who participated in a Pathfinder data collection methodology both before and after five days of DMO training. The Pathfinder methodology is a qualitative/quantitative method that can be used to assess if the pilots' underlying knowledge structures (i.e., their understanding) of air combat may have changed significantly as a function of DMO training. A total of 144 F-16 pilots rated the relatedness of 105 pairs of air combat concepts. Analyzing the before/after DMO training Pathfinder results by flight qualification or Viper flight position revealed remarkably similar results across all matrices. The most remarkable attribute of these networks is their stability; the pilots, regardless of demographic classification or before/after DMO training, view the relationships between the air combat concepts in very similar ways. We postulate that the stability of the networks most likely indicates that the 15 chosen concepts are at a high level of abstraction and reflect the shared knowledge in the general F-16 pilot population. As such, the concepts are likely not sensitive enough (i.e., detailed enough) to reflect differences in understanding resulting from five days of DMO training.

15. SUBJECT TERMS
Distributed Mission Operations; DMO; Flight simulation; Knowledge structures; MEC; Mission Essential Competencies; Pathfinder methodology; Pilot knowledge; Skill acquisition; Training assessment; Training effectiveness;

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr. Winston Bennett Jr
UNCLASSIFIED	UNCLASSIFIED	UNCLAS	UNLIMITED	30	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank

Table of Contents

List of Figures

List of Tables

EXECUTIVE SUMMARY

AFRL-HE-AZ-TR-2006-0015 – Vol I, Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Summary Report, describes a Distributed Mission Operations (DMO) training effectiveness study, examining 76 air combat teams (384 F-16 pilots). In that effort, three of the four general classes of datasets aimed to report direct measures of DMO within-simulator training effectiveness (i.e., objective data, subject matter expert observer judgments, and participant opinions). The fourth dataset, Pathfinder, was collected as an indirect measure of learning. The Pathfinder methodology is a qualitative/quantitative method that can be used to assess if the pilots' underlying knowledge structures (i.e., their understanding) of air combat may have changed significantly as a function of DMO training.

Using a 9-point scale, 144 F-16 pilots rated the relatedness of 105 pairs of air combat concepts. Participants were asked to perform the same relatedness judgments both before and after five days of air combat DMO training. The ratings among concept pairings are assumed to provide an estimate of the distance between concepts in memory. We expected that the Pathfinder tool, at the onset of DMO training, would reveal different knowledge structures as a function of F-16 experience level. As less experienced pilots learn more about the air-to-air combat concepts while flying in DMO exercises, results from post-DMO training concept ratings should reveal that their knowledge structures will become more stable and will reflect the permanence of the more expert knowledge structures.

The data were analyzed by occasion (before/after DMO training), flight qualification level (instructor, flight lead, or wingman), and F-16 cockpit assignment (Viper 1, Viper 2, Viper 3, or Viper 4). Across the six matrices formed by crossing occasion by qualification level of pilots, 14 links were in common across occasions and level of qualification, and the total number of links varied from a minimum of 16 to a maximum of 18. Similarly, for the 8 matrices formed by crossing occasion and Viper position, 15 links were common to all Viper positions across both occasions, and the total number of links varied between a minimum of 16 and a maximum of 18. Further, across all matrices, the concepts Linebacker for the Leading Edge and Clear Avenue of Fire emerged at the center of a stable set of nodes.

The most remarkable attribute of these networks is their stability. Whether classified by their level of qualification or the Viper position, most of the links are common across all levels of classification. Further, using either method, the same common links remained intact from the beginning to the end of the week of training. We believe that the stability of the networks most likely indicates that the 15 chosen concepts are at a high level of abstraction that reflect shared knowledge in the general F-16 pilot population, independent of receiving or not receiving specific training exercises such as the Mesa Research Site DMO. As such, DMO training may not affect the organization of these concepts.

ACKNOWLEDGEMENTS

The opinions expressed within this report are those of the authors and do not necessarily reflect the views of the Air Force or the sponsoring/employing organizations. This work was funded by the U.S. Air Force under contract #F41624-97-D-5000.

The authors wish to thank the following individuals for their significant contribution to this Appendix (in alphabetical order):

Michelle Borlase,
Jim Hartman,
Mary Johnson,
LtCol (ret.) Steve “Blaster” Masters,
LtCol (ret.) Michael “Odie” Park,
Dr Roger Schvaneveldt,
Traci Smith, and
Dr Brenda Wenzel.

This page intentionally left blank

DISTRIBUTED MISSION OPERATIONS WITHIN-SIMULATOR TRAINING EFFECTIVENESS BASELINE STUDY: USING THE PATHFINDER METHODOLOGY TO ASSESS PILOT KNOWLEDGE STRUCTURE CHANGES, VOLUME V

INTRODUCTION

AFRL-HE-AZ-TR-2006-0015 Vol I, Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Summary Report, describes a Distributed Mission Operations (DMO) training effectiveness study, examining 76 air combat teams (384 F-16 pilots). In that effort, Schreiber and Bennett (2006) reported numerous different data sources converging on the highly positive training effectiveness of an air combat DMO environment. As such, the report's focus was to document the overall results stemming from the central hypotheses of each dataset. Three of the four general classes of datasets aimed to report direct measures of DMO within-simulator training effectiveness (i.e., objective data, subject matter expert observer judgments, and participant opinions). The fourth dataset, Pathfinder, was collected as an indirect measure of learning. The Pathfinder methodology is a qualitative/quantitative method that can be used to assess if the pilots' underlying knowledge structures (i.e., their understanding) of air combat may have changed significantly as a function of the training. This report documents the Pathfinder knowledge structure assessment data in detail and provides us with the summary knowledge structure results used for the larger, summary report on DMO within-simulator training effectiveness (Schreiber & Bennett, 2006). Since the overall purpose of this report was to assess if knowledge structures were changing as a function of DMO training, it is important to emphasize that we used the Pathfinder methodology here as an assessment tool, and not to investigate esoteric academic questions concerning the Pathfinder methodology.

As part of the overall DMO training effectiveness study, F-16 pilots participated in a 5-day training research project, flying a total of 40 or more air combat scenarios. The overall demonstrated positive DMO within-simulator training effects are well documented in our other four volumes (Schreiber & Bennett, 2006; Schreiber, Stock, & Bennett, 2006; Schreiber, Schreiber, Gehr, & Bennett, 2006; Schreiber, Rowe, & Bennett, 2006). In the current work, we set out to use Pathfinder in an attempt to assess differences in the knowledge structures of participating DMO pilots. To make these knowledge structure assessments, the specific technique used in this research study for examining these changes is known as Pathfinder Network Scaling. Pathfinder has become increasingly popular as a means to assess knowledge structure, and it has been used previously to access knowledge structures among fighter pilots, computer programmers, students, and others (e.g., see Cooke & Schvaneveldt, 1988; Goldsmith, Johnson & Acton, 1991; Schvaneveldt et al., 1985; Schvaneveldt, Tucker, Castillo, & Bennett, 2001; Tessmer, Perrin, & Bennett, 1997). Commonly, the knowledge structure of experts and novices are compared/examined for differences in the way information is organized in memory.

In Pathfinder, these differences in how knowledge is organized in memory are assessed after having participants rate the degree of relatedness between pairs of concepts (e.g., in air-to-air combat, these concepts might include "build picture" and "make threat calls"). Participants rate the relatedness of many pairs of concepts, and they usually provide a range of relatedness

ratings, from low to high, depending on the concept pairing presented to them. These ratings among concept pairings are assumed to provide an estimate of the distance between concepts in memory. The Pathfinder algorithm generates a network (PFNET) based on mathematical graph theory (see Schvaneveldt, 1990; Schvaneveldt, Dearholt, & Durso, 1988). The concepts form the nodes on the graph and, where the concepts are related, the relationship between two nodes is represented by links. Weights are associated with these links and correspond to the strength of the relationship between two nodes. Pathfinder uses two parameters in computing these PFNETS. The q-parameter constrains the number of indirect proximities examined in order to generate the network. As q decreases the number of links added to the network can increase. For averaged proximity data, such as that to be used in the current work, it is recommended that $q = 2$ rather than the maximum $n-1$ (where n is the number of nodes or rating items). The r-parameter defines the metric used for computing the distance of paths. For ordinal data, the r-parameter is set to infinity.

To use the Pathfinder methodology as an assessment tool, we are assuming that current "expert" pilots (defined by high flight hours or flight qualification) do possess a stable and desirable knowledge network. From this assumption, we can then expect that the tool, at the onset of DMO training, would reveal different knowledge structures as a function of F-16 experience. As less experienced pilots learn more about the air-to-air combat concepts while flying in DMO exercises, results from post-DMO training concept ratings should reveal that their knowledge structures will become more stable and will reflect the permanence of the more expert knowledge structures (Shavelson, 1972).

Goals for this work included (a) to report the knowledge structure results in detail, and (b) to further augment the summary results as described in Volume I of the overall summary report of this training effectiveness study (Schreiber & Bennett, 2006). More specifically, we investigated the following hypotheses in support of the abovementioned goals:

1. Inexperienced pilot knowledge structures will be significantly different than expert pilot knowledge structures at the onset of DMO training.
2. Inexperienced pilot knowledge structures will exhibit increased similarity to the expert knowledge structure by the end of DMO training

METHODS

DMO Training

A portion of the following information is from General Method in Schreiber and Bennett (2006).

Table 1 shows a general timeline for each participating team. Teams were composed of four to six pilots and one AWACS operator. Participants arrived early Monday morning for five days of DMO participation. Upon arrival, participants were first given an inbrief on the objectives and procedures of DMO and the simulators, a tour of the facilities, and then given a research administrative session where they completed a demographic form, were assigned anonymous barcode identification numbers, and took the first Pathfinder exercise.

Table 1 Participant General Timeline

Session#	1	2	3	4	5	6	7	8	9
Day/time	Mon AM	Mon PM	Tues AM	Tues PM	Wed AM	Wed PM	Thur AM	Thur PM	Fri AM
Activity	Mesa Inbrief Admin Pathfinder Pilot Brief Fly Fam Pilot Debrief	Pilot Brief Fly 3 Benchs+ Pilot Debrief Feedback Survey	Pilot Brief Fly 4-8 engagements Pilot Debrief	Pilot Brief Fly 4-8 engagements Pilot Debrief	Pilot Brief Fly 4-8 engagements Pilot Debrief	Pilot Brief Fly 4-8 engagements Pilot Debrief	Pilot Brief Fly 4-8 engagements Pilot Debrief	Pilot Brief Fly 4-8 engagements Pilot Debrief	Pilot Brief Fly 3 Benchs+ Pilot Debrief Feedback Survey Reaction Survey Pathfinder Outbrief

Note: For the Monday/Friday benchmark assignment, pilots were asked to assign a flight position (i.e., Viper 1, 2, 3, or 4) to each pilot. These flight position assignments – the same for both Monday and Friday – were used in subsequent Pathfinder analyses.

Each scenario, Monday through Friday, was flown in the DMO training research environment in Mesa, AZ, which consisted of an instructor operator station, four high-fidelity F-16 simulators, computer-generated threats, and one high-fidelity Airborne Warning and Control System (AWACS) simulator. Four F-16 simulators, the AWACS, and threat entities interoperated according to Distributed Interactive Simulation (DIS) standards (IEEE Standard for Distributed Interactive Simulation - Application Protocols, 1995) version 4.02 or version 6.0.

The building-block training began immediately after the benchmarks (with the remaining time during session two) on Monday afternoon and continued through the course of the week. Each team was exposed to four to eight full engagements per session. While these training sessions emphasized Defensive Counter Air (DCA) scenarios, pilots also flew Offensive Counter Air (OCA) and air-to-ground missions. Participating teams experienced about 35 training engagements between the Monday and Friday benchmarks, providing an intensive training curriculum of over 40 total air combat engagements flown. The building block training sessions progressed in complexity by increasing the number of threat aircraft, the type of threat aircraft, the threat aircraft reactivity/maneuver, and/or an increase in the vulnerability time. Either after the last session on Thursday or on Friday morning, pilots took the second Pathfinder exercise and were given a DMO reaction rating form. Finally, before departure, teams were given a performance outbrief after their last set of benchmarks. This outbrief consisted of graphs for a number of the objective measures (Schreiber, Stock, & Bennett, 2006), revealing the team's performance.

Pathfinder Participants

All of the 384 pilots included in the larger DMO training effectiveness study were asked to complete the Pathfinder rating tasks. Of these pilots, 53 (13.8%) were missing either pre-

training or post-training Pathfinder data, or both. An additional 38 sets of pilot data were excluded either due to those pilots attending Mesa DMO (and completing Pathfinder) during prior years or they were part of an entire participating team whose data were eliminated. Finally, 149 pilots were eliminated for failing to meet a technical criterion for rating data coherence (described below). Overall, 144 pilots, only 37.6% of the original sample, were included in the Pathfinder analyses. Of these, all but two were male, with an average age of 32.3 years, an average of 9.9 years of service, and average number of 986.4 hours in an F-16. Broken out by flight qualification and flight position, the demographics were as follows:

Flight Qualification (three levels)

Instructor Pilot:	34.3 years old, 11.9 years of service, 1,385 F-16 hours
Flight Lead:	31.3 years old, 8.7 years of service, 712 F-16 hours
Wingmen:	28.7 years old, 7.0 years of service, 430 F-16 hours

Flight Position (four levels)

Viper 1:	32.3 years old, 9.7 years of service, 951 F-16 hours
Viper 2:	30.8 years old, 8.6 years of service, 702 F-16 hours
Viper 3:	33.2 years old, 10.8 years of service, 1,145 F-16 hours
Viper 4:	32.7 years old, 10.1 years of service, 943 F-16 hours

Demographics for the four flight positions (Vipers 1-4) reflect only those pilots who flew those positions during the Monday/Friday mirror-image benchmark scenario. Since teams brought four to six pilots to the DMO training, only four could fly in any given scenario, including the benchmarks. Therefore, demographics (and analyses) for the three flight qualification levels contain all 144 participants during the DMO training week, while the demographics (and analyses) for Vipers 1-4 contain only those 116 pilots that flew in the Monday/Friday mirror-image benchmark sessions. A quick review of the demographics reveals that many teams, for purposes of the DMO training, were likely intentionally assigning the most experienced pilots to other positions such as Viper 3 (element lead) or Viper 4 (Wingman), as opposed to the Viper 1 Flight Lead position, thereby allowing more inexperienced pilots opportunities at more flight responsibility.

Screening for Coherence

The before and after DMO distance matrices of each pilot was examined for coherence. As mentioned, 149 pilots' data were eliminated due to failing to meet coherence. Conceptually, coherence is a form of internal consistency. Schvaneveldt (1998; 2005) recommended eliminating matrices if the coherence value of the matrix is 0.20 or smaller. The coherence computation is based on the assumption that relatedness between a pair of items can be ascertained by the relatedness of each of the two items to the other items in the rating task. Coherence values less than 0.20 often indicate that raters are not taking the task seriously, and anecdotal reports from proctors of the Pathfinder data collection sessions indicate that this may have been an issue with a fair number of participants. The number of pilots excluded from analyses due to low coherence of before and/or after DMO training Pathfinder results was very high—38.8% of the entire 384 pilot sample. For the remaining 144 pilots whose data were used, the average coherence across before/after DMO training occasion was .41.

Pathfinder Rating Task

During the Pathfinder rating task, pilots were asked to use a 9-point scale to rate the degree of relatedness of 105 pairs of air combat concepts. These 105 pairs were the same in both the before and after DMO training Pathfinder sessions (sessions 1 and 9). The air combat concepts rated included both team members and tasks related to 2-ship employment with weapons director.

The endpoints and type of scale changed, depending on the pair of concepts rated, as described below. There were a total of 15 concepts, 3 concepts representing Team Member concepts, and 12 representing Task concepts. Each of the 15 concepts was paired once with each of the remaining concepts, producing the 105 unique concept pairs [(15*14)/2]. Note there were three different types of pairs formed, including; (a) Team Member – Team Member pairs (three total pairs), (b) Task – Task pairs (66 total pairs), and (c) Team Member – Task pairs (36 total pairs). The 3 Team Member and 12 Task concepts are given below.

<u>Concepts</u>

- *Team Member*
    - Flight lead
    - Wingman
    - Weapons Director
- *Task*
    - *S*anitize AOR (area of responsibility)
    - Make threat calls
    - Clear avenue of fire
    - Linebacker for leading edge
    - Listen
    - Mission flow
    - Build picture
    - Allocate radars
    - Formation/visual mutual support
    - Radar work to support shot(s)
    - Target as assigned
    - Weapons Engagement Zone (WEZ) denial

When a concept pair to be judged involved two Team Member concepts, a scale like the one depicted in Figure 1 was used. For each of these concept pairings, pilots based their rating on the degree to which the responsibilities of the two Team Members overlap. Essentially, the more overlap in responsibilities, the higher the rating.

```
 |-------|-------|------|-------|-------|-------|-------|------|
 1       2       3      4       5       6       7       8      9
None          "Overlap in Responsibilities"          Complete
```

Figure 1 Scale used when pilot was presented with a team member-to-team member concept pairing.

When a concept pair to be judged involved two Task concepts, a scale like the one shown in Figure 2 was used. For each of these pairs, pilots based their rating on the degree to which the outcome of one task directly affects accomplishing the other task. In general, the more impact a task has on the other, the higher the rating.

|-------|-------|------|-------|-------|-------|-------|------|
1 2 3 4 5 6 7 8 9
Not Interdependent "Task Relationship" Interdependent

Figure 2 Scale used when pilot was presented with a task-to-task concept pairing.

When a concept pair to be judged involved a Team Member concept and a Task concept, a scale like the one depicted in Figure 3 was used. For each of these pairs, pilots based their rating on the level of responsibility the named team member had for the named task. In general, the more responsible the team member was for the task, the higher the rating.

|-------|-------|------|-------|-------|-------|-------|------|
1 2 3 4 5 6 7 8 9
Not at All "Level of Responsibility" Fully

Figure 3 Scale used when pilot was presented with a team member concept and task concept pairing.

Choosing Levels of Aggregation

Following Schvaneveldt, et al. (1985) advice, we decided analyses of matrices of individual matrices of pilots involved a level of detail and degree of complexity beyond the scope of this investigation. Given the high number of matrices with low coherence values that were screened from any analyses, we judged that aggregated matrices offered the best opportunity to produce stable results. Consequently, we chose to aggregate the individual distance matrices in ways that were meaningful to an investigation of training effects on pilot performance. We implemented two different approaches to aggregation.

Our first approach crossed occasion (i.e., either before or after DMO training) with pilot qualification. Matrices were sorted into one of two groups representing Pathfinder ratings before and after DMO training. Thereafter, the set of "before" and the set of "after" DMO matrices were further divided into three subgroups corresponding to an Air Force qualification level of Instructor Pilot Qualification, Flight Lead Qualification, and Wingman Qualification. Thus, six sets of matrices were formed for aggregation by crossing two occasions with three levels of pilot qualification.

Our second approach crossed the before/after DMO training occasion with Viper position, i.e., Viper 1 (Lead pilot), Viper 2, Viper 3, and Viper 4. Thus, after separating matrices into groups corresponding to before and after DMO, matrices were further separated into groups representing Viper position 1 to Viper position 4. Hence, eight sets of matrices were formed for aggregation by crossing two levels of occasion with four Viper positions.

Aggregation

Each rating session produced a rating for each pair of the concepts, or 105 ratings. The ratings are averaged across the appropriate sets of data to create average ratings. These averages are used to derive the Pathfinder networks with the following general steps:

1. Invert the ratings so that the most highly related items will have the smallest values. A rating of 9 becomes 1, 8 becomes 2, and so on. The ratings then have the direction of distances.

2. Consider these data as representing a complete network with links between each pair of nodes. The weight of each link is the value of the inverted average rating for the pair of nodes connected by the link. Call these links the data links.

3. For all of the pairs of nodes, if the data link is a shortest path between the nodes connected by the link, retain the link, otherwise eliminate the link. With q = 2, the shortest path is either the direct link or an alternative path with two links. With r = infinity, the length of a two-link path is the maximum of the two link weights in the path.

Within each of the 14 sets of matrices formed (six from crossing occasion by qualification, and eight from crossing occasion by Viper position), an aggregate distance matrix was computed. The resulting networks form the primary findings of this study. For more detail regarding the generation of PFNETs, we refer the reader to Dearholt and Schvanevelt (1990).

RESULTS

Six networks were generated for the analyses by F-16 qualification, a “before” and “after” DMO training network for each of the three levels. Eight networks were generated for the analyses by Viper position (four positions each with a before and after DMO training network). These are described in separate sections below.

Before/After Occasion by F-16 Pilot Qualification

Figure 4 summarizes the weighted links of the networks derived from the three aggregate before DMO training matrices of Wingmen, Flight Leads, and Instructor Pilots, respectively. Figure 5 summarizes the links of the corresponding after DMO training matrices.

An examination of Figure 4 reveals that there is remarkable stability of the links among concepts for pilots at different levels of qualification on the before DMO training occasion. In fact, 14 links are common to all three levels of qualification (indicated by the black line connecting concepts). Further, the concept “Linebacker for the leading edge” is a node at the center of eight concepts – a cluster found for all three levels of qualification. A second cluster, centered by

“Clear avenue of fire” is nearly as large with five concepts linked to it. One of the latter five links is common to Instructor Pilots and Wingman, but not Flight Leads. Just two links (Wingman-Linebacker for the leading edge; Wingman-Mutual Support) are unique to a single level of qualification (Flight Lead).

An examination of Figure 5 reveals a nearly identical picture to Figure 4. First, the clusters of concepts are the same as those that appear in Figure 4. Second, the number of links shared by all three levels of qualification is high – 15, 14 of which also were shared links in Figure 4. In contrast to Figure 4, pilots with a Wingman qualification had three unique links in Figure 5 (Allocate Radar-Target as Assigned; Listen-Linebacker to the Leading Edge; Clear Avenue of Fire-Target as Assigned), but none in Figure 4.

Before/After Occasion by F-16 Pilot Viper Position

Figure 6 summarizes the weighted links of the networks derived from the four aggregate before DMO training matrices of Vipers 1 to 4, respectively. Figure 7 summarizes the links of the corresponding after DMO training matrices. An examination of Figures 6 and 7 reveal only small changes from the patterns observed in Figures 4 and 5. The same nodes centered around Linebacker for the Leading Edge and Clear Avenue of Fire are apparent in both Figures 6 and 7. In Figure 6, Before-DMO, there are two unique links, Wingman-Formation Visual Mutual Support for Viper Position 2 pilots; and Allocate Radars-Target as Assigned for Viper Position 4 pilots), whereas in Figure 7, after DMO training, there were three unique links (Build Picture-Weapons Director and Flight Lead-Radar Work to Support Shot for Viper Position 3 pilots; and Wingman-Linebacker for Leading Edge for Viper Position 4 pilots).

In sum, across the six matrices formed by crossing occasion by qualification level of pilots, 14 links were in common across occasions and level of qualification, and the total number of links varied from a minimum of 16 to a maximum of 18. Similarly, for the eight matrices formed by crossing occasion and viper position, 15 links were common to all viper positions across both occasions, and the total number of links varied between a minimum of 16 and a maximum of 18. Further, across all matrices, the concepts Linebacker for the Leading Edge and Clear Avenue of Fire emerged at the center of a stable set of nodes.

When examining the networks displayed in Figures 4 to 7, we remind the reader that the links in each of the figures simply represent that a link exists. Distances between nodes, and positioning of links from node to node are arbitrary. The authors constructed the arrangements to visually facilitate understanding of the networks of concepts, and the stability that these networks display, a phenomenon that the authors see in each individual network, as well as across the set of four networks.

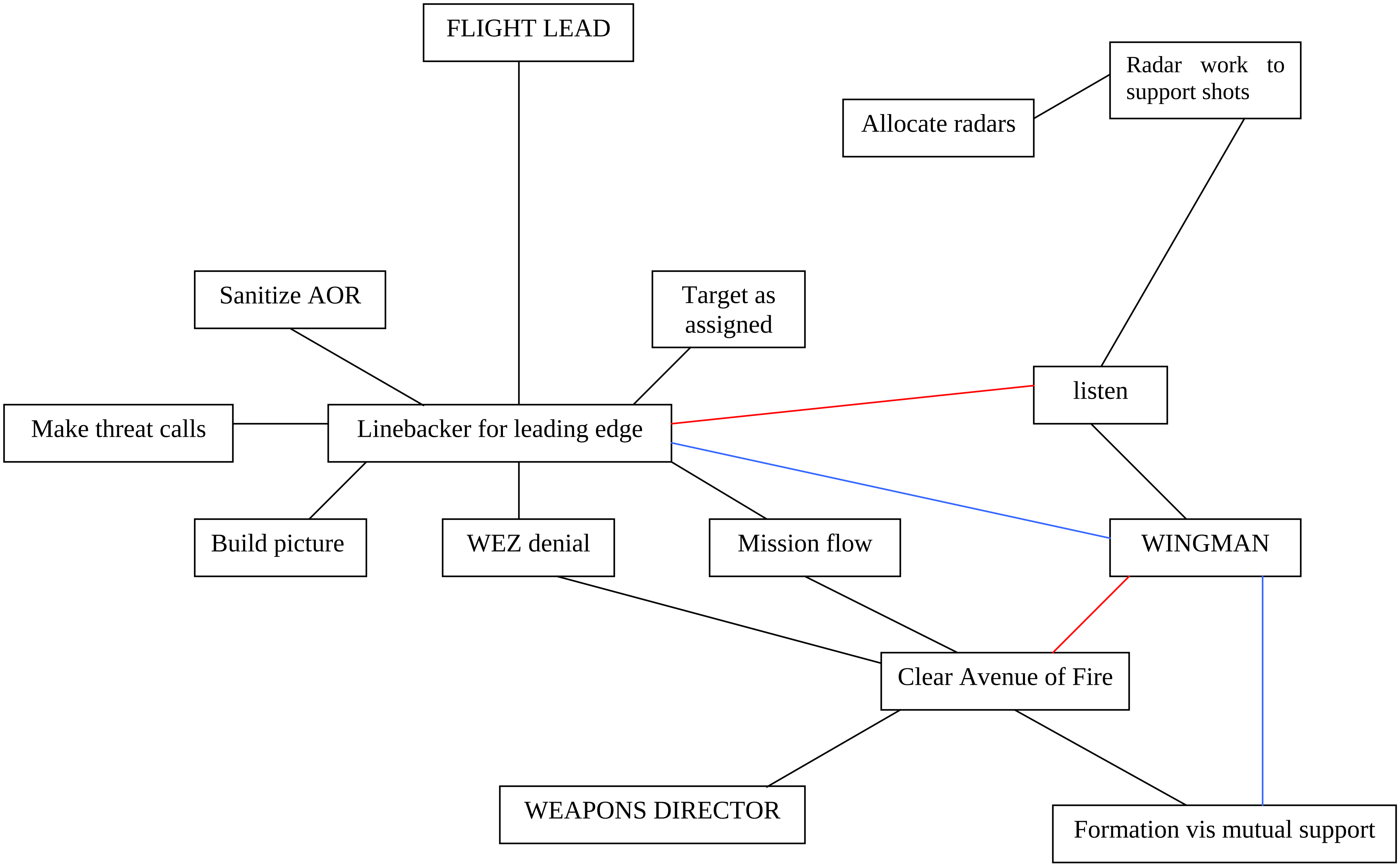

Figure 4 Before DMO Training PFNET Differences by F-16 Qualification Level

— Link common to all networks — AvgLeaPre Only — AvgIPpre & AvgWingPre Only

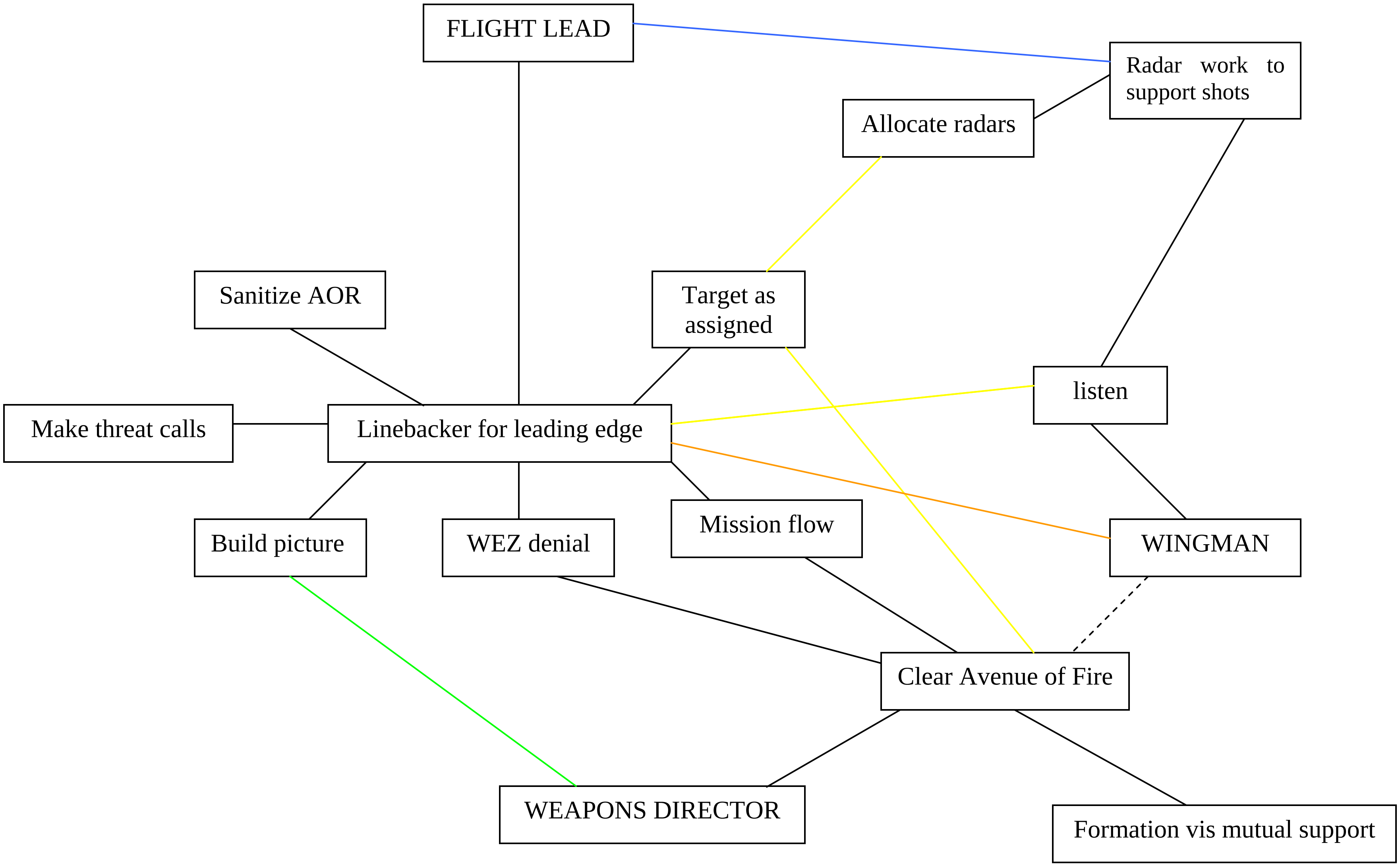

Figure 5 After DMO Training PFNET Differences by F-16 Qualification Level

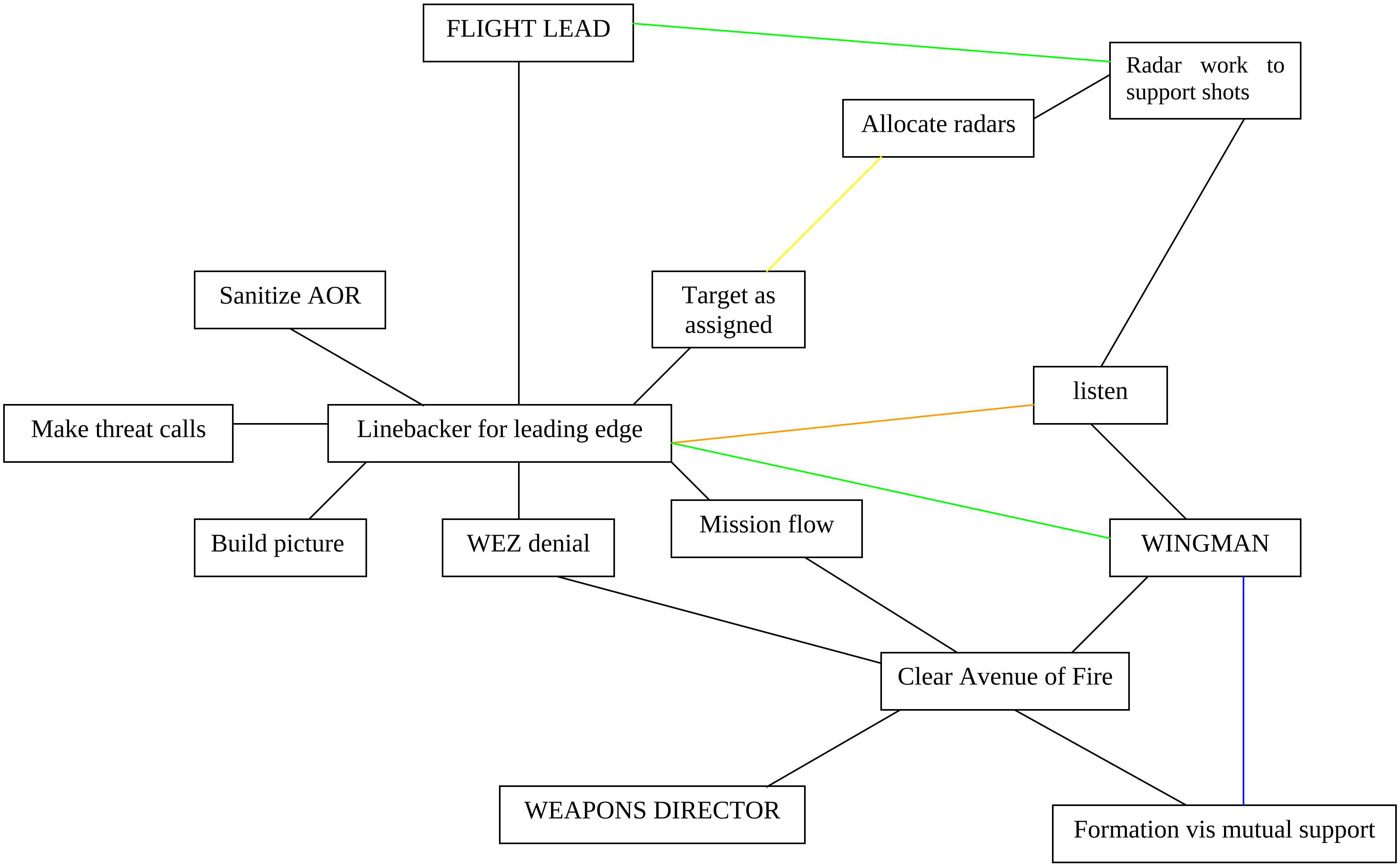

Figure 6 Before DMO Training PFNET Differences by Benchmark Viper Position

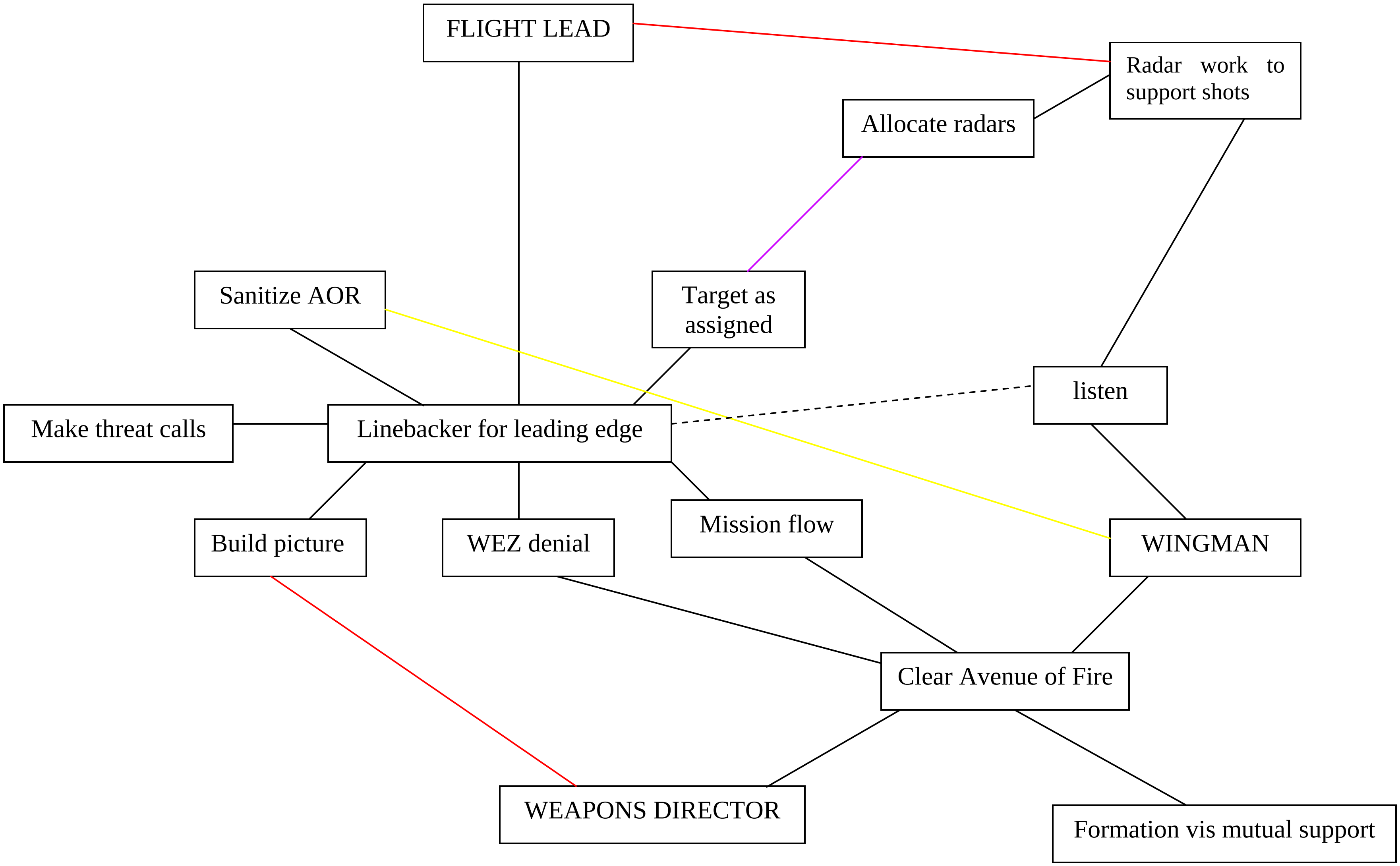

Figure 7 After DMO Training PFNET Differences by Benchmark Viper Position

— Link common to all networks
--- Link common to all Post networks
— AvgV3Post Only
— AvgV2Post & AvgV4Post Only
— AvgV4Post Only

DISCUSSION

Unfortunately, a significant result of the current work was that we obtained useable data from less than half of the original sample. Of the original 384 pilots, 53 (13.8%) were missing either pre-training or post-training Pathfinder data, or both. Additionally, and more troubling, 149 pilots were eliminated for failing to meet a technical criterion for rating data coherence. Because we have found that fighter pilots nearly universally commit and comply to training research protocols, the fact that 13.8% of pilots failed to complete some aspect of the Pathfinder task raises some concerns about the pilots' willingness to complete the task. Strengthening that concern, the extremely high proportion (38.8%) of pilots that did not meet coherence raises great doubt about the degree to which the pilots' felt the task possessed ecological validity. For the 144 pilots whose data were used in the current work, we assume that this was not an issue, or at least not sufficiently so to adversely affect their data.

The most remarkable result of the remaining reported networks is their stability. Whether the pilots are classified by their level of qualification (Figures 4 and 5) or the Viper position that they flew during the Monday and Friday benchmarks (Figures 6 and 7), most of the links are common across all levels of classification. Further, using either method of classification of pilots, the same common links remained intact from the beginning to end of the week of training. An attribute of all of the networks was that the concept "Linebacker for the leading edge" was a node with the most paths to it, and that the concept "Clear Avenue of Fire" was a node with the next highest number of paths to it. A third concept, "Radar work to support shots," represented a third stable cluster, but with the fewest paths connecting to it.

In independent sessions, we separately asked three subject matter experts (SMEs), all recently retired F-16 pilots, to complete the following post-hoc assessment task of the Pathfinder networks shown in Figures 4 through 7. First, we showed the SME Figures 4 and 5, starting with Figure 4, and explained the time of measurement, the meaning of the links (black line indicates all three levels of qualification [Wingman, Flight Lead, Instructor Pilot] had the same link), and specifically noted those links where not all levels of qualifications had the same link. In Figures 4 and 5, there are four and six such cases, respectively (In Figures 6 and 7, there are five and four such cases.). At that point, we asked each SME to provide a principled rationale for the observed differences pre-training, and for the changes from pre- to post-training. This process was repeated for Figures 6 and 7.

For both pairs of network depiction (Figures 4 and 5, Figures 6 and 7), the first SME simply stated that he was unable to construct a rationale for the observed pattern in the target links (i.e., those not shared across all levels of qualification). The second SME suggested that instructor pilots are generally more oriented toward the big picture, and at the beginning of the week have a perspective more like that of a Wingman, rather than a Flight Lead, who is concentrated on the immediate tasks of the mission. The minor changes from pre- to post-training he attributed to learning. The third SME essentially reviewed each of the target links, and indicated with which he agreed or disagreed. Thus, in Figure 4, he indicated agreement for the "Linebacker for the leading edge-listen," "Wingman-Clear Avenue of Fire," and "Wingman-Formation visual support," but not "Linebacker for the leading edge-Wingman," and repeated the same judgments in Figure 5 for "Linebacker for the leading edge-listen" (agree), and "Linebacker for the leading

edge-Wingman" (disagree), as well as for "Linebacker for the leading edge-Wingman" in Figure 6 (disagree), and "Linebacker for the leading edge-listen" in Figure 7 (agree). He reported that he based his judgments on the rationale that the Wingman's role is to support the effort, i.e., be there in such a way that the Flight Lead can work unencumbered. Clearly, the SMEs did not converge on similar explanations for the minor differences in results, but they did seem to generally indicate that many of the links were understandable. This could be the result of the concepts containing higher level abstractions, and the networks revealing understanding common across F-16 pilots. The minor differences observed could simply be attributable to noise, especially given the fairly low coherence values.

The Pathfinder networks are more similar than dissimilar across time, and across two different levels of classifying pilot expertise. To use the Pathfinder methodology as an assessment tool, we assume that expert pilots do possess a stable and desirable knowledge network. Across the networks in Figures 4 and 5, where there are small differences, these differences do not align neatly with the classification of pilots by qualification. From Figure 6 to 7, there is a tendency for Vipers 2 and 4 to have paths in their networks that are not present in the networks of Vipers 1 and 3. This finding is mitigated by the fact that many of the Viper 1 pilots were less experienced pilots placed in the Viper 1 position to gain valuable experience as a Flight Lead. Hence, we are left with the impression that it is the stability of the networks that is remarkable here.

The DMO training week is intended to alter the behavior of pilots by engaging them in meaningful and challenging tasks, and by giving them sufficient practice to master new skills, become more competent warfighters, and come away with a stronger grasp of the organizing constructs of air combat. As clearly shown in other volumes of this technical report series (Schreiber, Gehr, & Bennett, 2006; Schreiber, Rowe, & Bennett, 2006; Schreiber, Stock, & Bennett, 2006), the forms of direct assessment data convincingly demonstrate this outcome. Hence, we must reconcile the findings of the indirect assessment here with the findings of the other volumes.

It is the authors' opinion that the great stability in the networks most likely indicates that the 15 chosen concepts are at a high level of abstraction that reflect shared knowledge in the general F-16 pilot population, independent of receiving or not receiving specific training exercises like Mesa DMO. As such, training may not affect the organization of these concepts at all. This is even more probable given the demographic. Even the "low experienced" pilots still had seven years of service and 430 flight hours; a ceiling effect could be underlying the lack of differences. Especially with an experienced population, the chosen concepts may not be close enough to the experiences gained in the DMO exercises (i.e., sensitive enough) to reflect the cognitive changes that these experienced pilots do report occurring (Schreiber, Rowe, & Bennett, 2006). We recommend further refinement of the concepts at a lower level of abstraction if the Pathfinder task is to continue to be used as an indirect measure of learning in the DMO environment.

REFERENCES

Cooke, N. J., & Schvaneveldt, R. W. (1988). Effects of computer programming experience on network representations of abstract programming concepts. *International Journal of Man-Machine Studies*, *29 (4*), 407-427.

Dearholt, D. W., & Schvaneveldt, R. W. (1990). Properties of Pathfinder networks. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization*. Norwood, NJ: Ablex Publishing Co.

Goldsmith, T., Johnson, P., & Acton, W. (1991). Assessing structural knowledge. J*ournal of Educational Psychology, 83(1),* 88-96.

Schreiber, B. T. & Bennett, W. Jr. (2006). D*istributed Mission Operations within-simulator training effectiveness baseline study: Summary report.* (AFRL-HE-AZ-TR-2006-0015-Vol I). Mesa AZ: Air Force Research Laboratory, Warfighter Readiness Research Division.

Schreiber, B. T., Gehr, S. E., & Bennett, W. Jr. (2006). *Distributed Mission Operations within-simulator training effectiveness baseline study: Real-time and blind expert subjective assessments of learning.* (AFRL-HE-AZ-TR-2006-0015-Vol III). Mesa AZ: Air Force Research Laboratory, Warfighter Readiness Research Division.

Schreiber, B. T., Rowe, L. J., & Bennett, W. Jr. (2006). *Distributed Mission Operations within-simulator training effectiveness baseline study: Participant utility and effectiveness opinions and ratings.* (AFRL-HE-AZ-TR-2006-0015-Vol IV). Mesa AZ: Air Force Research Laboratory, Warfighter Readiness Research Division.

Schreiber, B. T. & Bennett, W. Jr. (2006). *Distributed Mission Operations within-simulator training effectiveness baseline study: Metric development and objectively quantifying the degree of learning.* (AFRL-HE-AZ-TR-2006-0015-Vol II). Mesa AZ: Air Force Research Laboratory, Warfighter Readiness Research Division.

Schvaneveldt, R. W. (Ed.) (1990). *Pathfinder associative networks: Studies in knowledge organization*. Norwood, NJ: Ablex Publishing Co.

Schvaneveldt, R. W. (1998). Info.doc. [Computer program documentation file]. Gilbert, AZ: Interlink.

Schvaneveldt, R. W. (2005). Personal Communication. Arizona State University Polytechnic Campus, Mesa, AZ.

Schvaneveldt, R. W., Dearholt, D. W., & Durso, F. T. (1988). Graph theoretic foundations of Pathfinder networks. *Computers and Mathematics with Applications, 15*, 337-345.

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., Cooke, N. J., Tucker, R. G., & DeMaio, J. C. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies*, *23,* 699-728.

Schvaneveldt, R., Tucker, R., Castillo, A. R, & Bennett, W., Jr. (2001). Knowledge acquisition in distributed mission training. In *2001 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Shavelson, R. J. (1972). Some aspects of the correspondence between content structure and cognitive structure in physics instruction. *Journal of Educational Psychology*, **63**, 225-234.

Tessmer, M., Perrin, B., & Bennett, W., Jr. (1997). *Assessing the stability of structural learning measures* (AL/HR-TR-1997-0131, AD368474). Brooks AFB TX: Human Resources Directorate, Cognition and Performance Division.

ACRONYMS

AFRL	Air Force Research Laboratory
AOR	Area of Responsibility
AWACS	Airborne Warning and Control System
DCA	Defensive Counter Air
DIS	Distributed Interactive Simulation
DMO	Distributed Mission Operations
IEEE	Institute of Electrical and Electronics Engineers
OCA	Offensive Counter Air
PFNET	Pathfinder Network
SME	Subject Matter Expert
USAF	United States Air Force
WEZ	Weapons Employment Zone

This page intentionally left blank

APPENDIX A: Instruction Sheet Verbally Presented to Participants

The following contains the appropriate script and procedures for conducting Pathfinder data collection sessions.

WELCOME: Hello, I am (Use your name here). I am one of the AFRL Mesa team members you will see quite a bit over the next week. I'd like to take this opportunity to again welcome you to the Distributed Mission Training research testbed. Over the course of the next week, you will get a chance to participate in some very comprehensive and challenging combat mission training simulation research.

As part of this research, a number of us will be asking you to provide information throughout the week. The information will typically be your reactions and specific feedback to the missions and engagements you have experienced. You will also be asked to participate in completing questionnaires and other activities that will help us determine the usefulness of the missions and engagements for improving learning and air combat mission readiness.

This session represents one of the first opportunities for you to provide us with some information of this type.

General Information to be provided to participants prior to starting Pathfinder:

You are about to participate in an ongoing research study to assess how you think about the interrelationship of key air combat mission concepts. For our purposes, a concept refers to a given aspect of air combat. You'll see that the concepts are very familiar to you as a fighter pilot.

The way you think about these concepts and their relationship to other concepts helps us in understanding the impact of the training research missions you will fly (or have flown) this week. Since a considerable amount of our research involves trying to impact this understanding and the learning that goes along with combat mission training, the information you are providing in this session is extremely important. Your information will help us identify ways to improve the training and rehearsal activities that you and future pilots will receive. We have tried to keep the data collection process as simple as possible by limiting the number of concepts you will see and by automating it. You will be using a software program called "Pathfinder" which has been installed on each of the machines in this room.

You will be asked to complete the Pathfinder program twice during your exercise. The first will be at the beginning of the week, before you begin the simulation training activities. The final session will be after you have completed your last simulation training mission here at Mesa on Friday.

Pathfinder contains 3 team member concepts and 12 task concepts that relate to air-to-ground combat and to the environment you will experience while you are here. You will be presented with two of the concepts at a time and will be asked to judge how closely the two

concepts are related to one another on three types of scales. The scale and the endpoints of the scale are labeled differently based on the type of pair being rated. Examples of the different scale labels that accompany the different pair types (2 tasks, 2 team members, team members and task) are shown on your instruction sheet.

Your response will be based on a 9-point scale where 1 indicates that the concepts are not related and a rating of 9 indicates the highest degree of relatedness. You will continue this rating process for all the concept combinations. The entire rating process should take you no more than 30 minutes, so please be patient and answer as best you can. You should not spend a lot of time on each comparison, but simply give the response that best represents your initial impression.

This is not a graded or a timed activity and there are no “right” or “wrong” answers. Remember, this is being done to study overall thought processes and training effectiveness and is not meant to measure individual performance.

Please do not ask questions or discuss the concepts with each other during the session. If you have any questions or technical problems during the exercise, please ask me for assistance. Thank you for participating in this part of the process and thank you for assisting us in the data collection effort. I hope your experience here at Mesa is productive and meaningful. I’ll see you many times over the next week, so if you have any questions that pop up, please ask me or the other Mesa team members. We’re here to help.

APPENDIX B: Participant Reference Instructions

We are interested in the way that you think about your team responsibilities/tasks/duties, responsibilities/tasks/duties of other members in your team, and how those responsibilities/tasks/duties interrelate.

<u>Assumptions</u>:

1. Enemy Air-to-Air threat capability = Medium
2. Enemy Surface-to-Air threat capability = Medium
3. Weapons Director on frequency
4. Enemy communications jamming threat = Low

<u>Scenario</u>:

You are in a four-ship which is part of a larger strike package. The strike package is fragged against a large enemy HQ complex approximately 75 nm behind the FEBA. Your target is a small Power Generation Building in the NW corner of the complex.

Here is the list of items that you will be rating in pairs:

<u>Team Members</u>: FLIGHT LEAD WINGMAN WEAPONS DIRECTOR

<u>Tasks</u>:

listen	Close Air Support	medium level attack
formation/vis mutual support	defend against surface threats	low level attack
Combat ID	Assess attack validity against ROE	Time Sensitive Targeting
Make threat calls	ECM employment	Target as assigned

Each pair will be presented on the screen along with a "relatedness" scale. The scale and the endpoints of the scale are labeled differently based on the type of pair being rated. Examples of the different scale labels that accompany the different pair types (2 tasks, 2 team members, team members and task) are shown below

```
|-----------|-------------|-----------|----------|-----------|----------|----------|-----------|
1           2             3           4          5           6          7          8           9
```

NOT INTERDEPENDENT
(Outcome of one task does NOT directly affect the other tasks)

"Task Responsibility"
(2 tasks)

INTERDEPENDENT
(Outcome of one task does directly affect the other task)

```
|-----------|-------------|-----------|----------|-----------|----------|----------|-----------|
1           2             3           4          5           6          7          8           9
```

NONE

"Overlap in Responsibilities"
(2 team members)

COMPLETE

```
|-----------|-------------|-----------|----------|-----------|----------|----------|-----------|
1           2             3           4          5           6          7          8           9
```

NOT AT ALL

"Level of Responsibility"
(team member and task)

FULLY

The degree to which you judge the two items in the pair to be unrelated (Not Interdependent, None) or related (Interdependent, Complete, Full) should determine the rating between 1-9 that you give the pair of items.

Indicate your rating by pressing a number key on the keyboard. Upon responding, a bar marker will move directly above the number you pressed. If you wish to change your response, simply press another number. When you are satisfied with the rating you have given, press the SPACE BAR to enter your response. The next pair of items to be rated will then be displayed. The list of items to be rated will be presented on the final screen of the computer presented instructions.